

Lustre usage by DiRAC

UK Lustre User Group

Manchester

Dec 2022

Alastair Basden

Durham University / DiRAC



Durham
University

DiRAC
High Performance
Computing Facility

DiRAC introduction

- STFC-funded UK Tier-1 HPC facility
 - HPC systems at 4 sites:
 - Cambridge
 - Durham
 - Edinburgh
 - Leicester
 - Several systems in TOP500

Lustre @ DiRAC

- All systems use Lustre
- Cambridge:
 - 8 Lustre systems, total of 4.8PB
- Durham:
 - 7 Lustre systems: 0.7PB, 1.6PB, 2.6PB, 3.1PB, 7PB (~16PB total)
 - Including 2 fast scratch systems, 460TB and 1.2PB (~400GB/s)
- Edinburgh:
 - 2 Lustre systems: 3PB, 4PB, DDN appliances
- Leicester:
 - 2 Lustre systems, HPE (now community) 3.4PB, Cray ClusterStore 3.8PB

Lustre@Durham

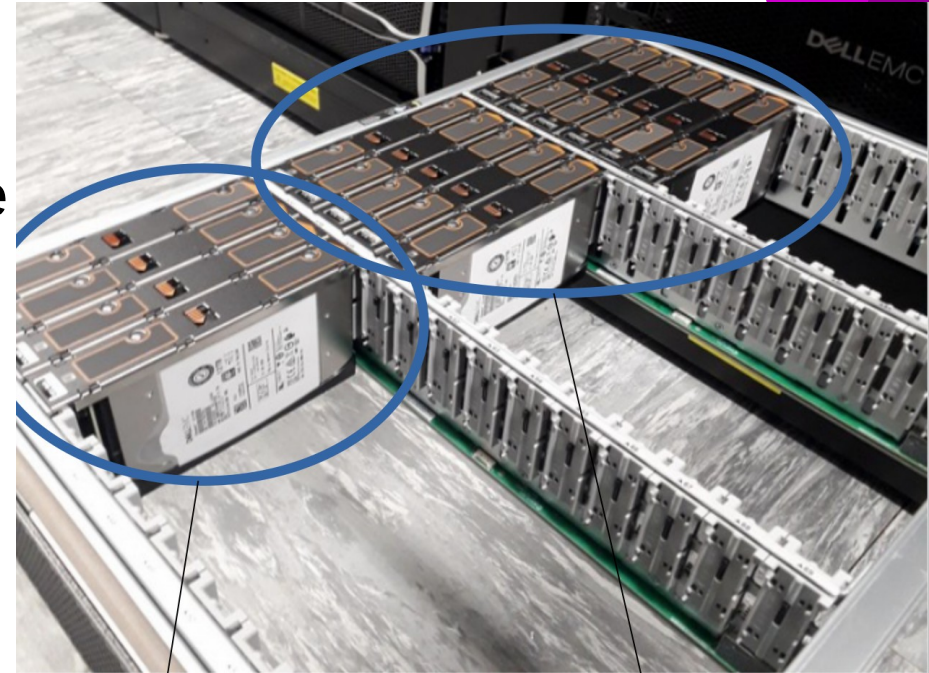
- /madfs (0.7PB): 1 MDS, 1 MDT, 8 OSS, 12 OST - ldiskfs/ldiskfs (controller)
- /cosma5 (1.6PB): 1 MDS, 1 MDT, 2 OSS (1 shared), 14 OST - zfs/zfs
- /cosma6 (2.6PB): 2 MDS, 2 MDT, 4 OSS, 28 OST - ldiskfs/zfs
- /cosma7 (3.1PB): 1 MDS, 1 MDT, 16 OSS, 54 OST - ldiskfs/ldiskfs (controller)
- /cosma8 (7PB): 4 MDS, 4 MDT, 8 OSS, 56 OST - ldiskfs/zfs
- /snap7 (460TB): 1 MDS, 1 MDT, 20 OSS, 160 OST - ldiskfs/ldiskfs
- /snap8 (1.1PB): 1 MDS, 4 MDT, 24 OSS, 192 OST - ldiskfs/ldiskfs
- All on Dell hardware

Our experience

- Lustre is *mostly* stable
 - Powerful
- Some features could be improved
 - PFL implementation
 - Hanging occurs reasonably often with PFL
 - Slow response times
 - Other bugs occasionally surface
 - Mirroring
 - Relies on user to mirror
 - Both copies can be on same disk if pools aren't used
- Some of the tools appear forgotten, e.g. IML
- Some of the documentation is unclear or outdated
- Seems to know when it is Friday evening or Christmas Eve

Setup

- In theory, reasonably straightforward
 - But lots of fine tunings can be made
 - Default options not always most sensible
- Significant effort spent optimising and balancing SAS channels
 - Now have a script to do this



Problems

- A number of current problems
 - None are showstoppers
 - Some are inconveniences
 - Some are annoyances

Mounting MDTs

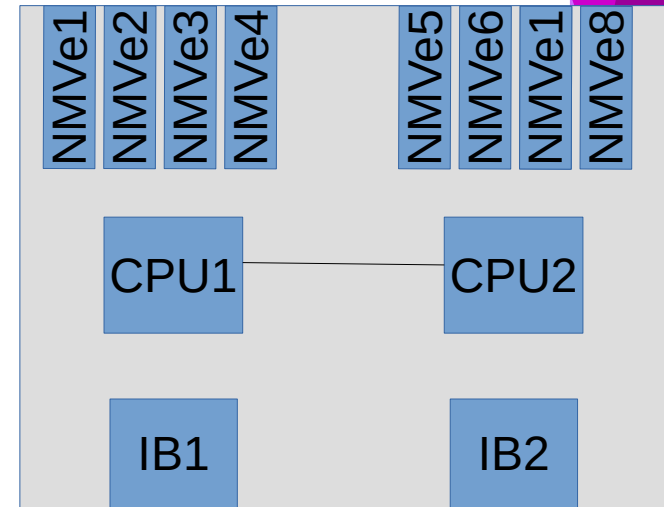
- Mounting the MDTs hangs until the OSTs connect (or 25 minute timeout is reached)
 - New feature by design in 2.12?
 - Recovery window
 - Problematic if a OST (or OSS?) is unavailable

Quotas

- Occasionally seem to be ignored
 - i.e. a user can go significantly over quota and still be able to write

Dual IB routing

- Fast NVMe storage servers with 2 IB connections
 - 4 NVMe drives served by each CPU using a dedicated IB connection
 - Writing to groups of 4 disks seems to fail at times, after a fresh server boot
 - Those OSTs get marked as down by the clients
 - Not always the same 4, but usually on the second CPU of a server
 - Always on one half of the system, sharing IB link.
 - Reboots of the servers eventually fix this. Something to do with Inet routing, we think.



Speed

- Occasional incredibly slow file access, when load on OSSs goes very high. Possibly due to underlying OST contention.
 - But I/O rates to disk seem low
 - Probably nothing can be done - except direct new writes to alternative servers while load is high
 - PFL doesn't help: significantly increases file system load

Data loss

- November 2021
- Pacemaker failover led to corrupted ZFS OST
 - This shouldn't be possible
 - Bug in ZFS 0.7.9

zfs MDT kernel panics

- kernel: VERIFY(dnode_add_ref(dn, (void *) (uintptr_t)tx->tx_txg)) failed
 - kernel: PANIC at dnode.c:1635:dnode_setdirty()
 - Known bug, backport fix for ZFS not yet done
 - Very infrequent (~twice)

Suspended pools

- Zpools being suspended due to:
 - “the pool is suspended because multihost writes failed or were delayed”
 - Lack of integration between Lustre and ZFS
 - Increasing `zfs_multihost_interval` larger reduces this problem
 - ZFS multimount protection
 - Automatic failover seems risky

>2.12.6

- Current situation is deterministic instability for server/client versions newer than 2.12.6
 - MPI jobs get stuck if they run out of memory, tidying up fails, InfiniBand gets upset, reboot required.
 - Only a client issue

Nids, mounts etc

- Re-mounting over the wrong interface
 - Occasionally, the IB mounts on clients to some OSTs are replaced by TCP mounts.
 - `lctl dl -t | grep tcp`
 - Requires umount of the file system on client to fix
 - We think due to temporary unresponsiveness causing the client to think that NID has failed
 - Unclear how to change without remounting

SIGBUS

- Codes see occasional SIGBUS errors on compressed ZFS backends

Summary

- Lots of Lustre systems across DiRAC
 - Lot of experience within DiRAC teams
- Mostly works well
 - But there are niggles